



ELSEVIER

Discrete Applied Mathematics 88 (1998) 3–6

**DISCRETE
APPLIED
MATHEMATICS**

Foreword

This is the second special volume of *Discrete Applied Mathematics* focusing on the development of new combinatorial and algorithmic techniques in computational molecular biology. The first special volume was published in 1996 as Volume 71 of *Discrete Applied Mathematics*. The Journal will continue to publish regularly volumes on computational biology as part of the recently initiated *Computational Molecular Biology Series (DAM–CMB Series)*. As Series Editors, we will continue to assure a thorough and timely refereeing process that will follow the general pattern for refereeing of regular papers in the journal.

We took as “motto” for our series, the comments made by Stanislaw Ulam in 1976 about the richness of mathematical research that will be inspired by Biology.

“After reading about the new discoveries in molecular biology which were coming fast, I become curious about a conceptual role which mathematical ideas could play in biology. If I may paraphrase one of the President Kennedy’s famous statements, I was interested in “not what mathematics can do for biology but what biology can do for mathematics”. I believe that new mathematical schemata, new systems of axioms, certainly new systems of mathematical structures will be suggested by the study of the living world. Its combinatorial arrangements may lead us in the future to a logic and mathematics of a different nature from what we know now.”

The second volume of the DAM–CMB Series contains a diversity of papers demonstrating again the richness of this new and exciting area. We have taken the liberty of partitioning the papers into five areas: mapping and sequencing, molecular evolution, genome rearrangements, RNA and protein structure, and computational support for experiments.

Mapping and Sequencing. The papers in this volume focus on new algorithms for the consecutive ones property, interval graphs, superstrings, and Hamiltonian cycles problems relevant to DNA mapping and sequencing. In the paper “Graph traversals, genes and matroids: an efficient case of the traveling salesman problem” by D. Gusfield, R.M. Karp, L. Wang and P. Stelling, graph traversal problems that are motivated by sequencing by hybridization (SBH) technology are considered. Exact and approximation algorithms are presented for these SBH problems, including a practical polynomial time solution to the traveling salesman problem for a rich class of directed graphs. This is the first computational biology paper providing connections with matroid theory. In the papers “On testing Consecutive Ones Property in parallel” by F. Annexstein and

R. Swaminathan, and “On the Consecutive Ones Property” by J. Miedanis, O. Porto and G.P. Telles, the authors study the consecutive ones property, presenting algorithms based on new data structures. The first presents new insights and data structures for algorithms with improved performance for parallel computations, while the second presents a generalized PQ-tree data structure that can handle arbitrary collections of sets. In the paper “On probe interval graphs” F.R. McMorris, C. Wang and P. Zhang study a family of graphs that is a generalization of interval graphs. These graphs arise in physical mapping of DNA, where only the intersections of a subset of the intervals with all intervals are known. The paper presents characterizations and recognition algorithms for the family. Grebinsky and G. Kucherov in the paper “Reconstructing a Hamiltonian circuit by querying the graph: application to DNA physical mapping” study the mathematical model of multiplex PCR for chromosome mapping. The problem is cast as combinatorial group testing, leading to efficient algorithms matching information-theoretic lower bound. The paper “A 2 and $\frac{3}{4}$ approximation algorithm for the shortest superstring problem” by C. Armen, C. Stein, uses new structural theory on periodicity in strings in order to obtain an improved approximation algorithm for the shortest superstring problem.

Molecular evolution. The papers in this volume are devoted to stochastic models, phylogenetic trees, and tree alignment. The paper by O. Eulenstein and M. Vingron entitled “On the equivalence of two tree mapping measures” provides the solution to a conjecture related to the mathematical difficulty of reconciling different evolutionary trees. In the paper “On the approximability of the Steiner tree problem in phylogeny” by D. Fernandez-Baca and J. Lagergren, several complexity and algorithmic results are obtained for the problem of computing phylogenies with a specified number of reconstructed ancestral sequences. “Approximation algorithms for multiple sequence alignment under a fixed evolutionary tree” by R. Ravi and J. Kececioglu, describes approximation algorithms for several new objective functions together with varying degrees of provability of the algorithms performances. J. Kececioglu and D. Gusfield in the paper “Reconstructing a history of recombinations from a set of sequences” consider the problem of reconstructing an evolutionary history of species evolving by recombination. The authors develop the algorithms for recombination cost and bottleneck recombination history problems motivated by non-tree models of evolution. The paper “Reconstructing phylogenies from nucleotide pattern probabilities” by M. Steel, M.D. Hendy and D. Penny is a survey containing also new results on Markov-model based phylogenetic tree reconstruction from expected frequencies of patterns in the sequences.

Genome Rearrangements. Genomic rearrangements is a new and very active research area. Two papers focusing on rearrangement algorithms and a new genomic distance

measure are included. Algorithmics related to genomic rearrangements by switching adjacent blocks are presented in the paper “Sorting by bounded block-moves” by L. Heath and J. Vergara. In the paper “On the complexity of approximation of syntenic distance” B. DasGupta, T. Jiang, S. Kannan, M. Li, Z. Sweedyk study the syntenic distance: a new rearrangements-based distance measure between multi-chromosomal genomes.

Protein and RNA Structure. This fundamental area of research needs new mathematical methods to provide computational tools for studying protein and RNA structure prediction. Two papers in this volume are devoted to the geometry of protein complexes and the combinatorial enumeration of RNA folds. A paper dealing with computational geometry problems related to protein structure is “On the definition and the construction of pockets in macromolecules” by H. Edelsbrunner, M. Facello and J. Liang. It presents an algorithm for locating “pockets” in protein complexes; they capture the notion of region in the complement space (defined by protein interactions) that has limited accessibility from the outside (the protein complex). Secondary structures of polynucleotides can be viewed as a class of planar vertex-labeled graphs. I. Hofacker, P. Schuster and P. Stadler, in the paper “Combinatorics of RNA secondary structures” present recursion formulae for enumerating a variety of sub-classes and classes of subgraphs (structural elements) of secondary structure graphs.

Computational Support for Laboratory Experiments. An important component of the experimental methods, used for the compilation of physical maps of chromosomes, and other genetic material, is the pooling methodology. In the paper “Non-adaptive group testing in the presence of errors”, E. Knill, W. Bruno and D. Torney present a unified treatment of the known methods for constructing pooling designs, provide explicit formulas for their performance under different error assumptions, and discuss the asymptotic performance of random designs. Branch and Bound algorithms are presented in the paper “Genotyping of pooled microsatellite markers by combinatorial optimization techniques” by G. Lancia, M. Perlin in order to provide computational support for genotyping. Using these methods, genotyping of pooled markers can be computed effectively, thus potentially achieving a considerable reduction in experimental time and expense.

We would like to thank to Dr. Peter Hammer, the Editor-in-Chief of *Discrete Applied Mathematics* for his enthusiastic support of this series, and his advice and guidance on all the aspects of the editorial process. We would like to express our appreciation to Mrs. Nelly Segal, the Editorial Manager of the journal, who helped us in many ways with the management and the final preparation of the volumes. We also want to thank the anonymous referees for their hard work that helped us in evaluating the submissions for this volume.

We hope that this series will continue to publish top quality research papers that will advance the state of the art of the exciting new area of computational biology.

Sorin Istrail
Pavel Pevzner
Ron Shamir
Guest Editors

Sorin Istrail
Algorithms and Discrete Mathematics Department
Sandia National Laboratories
P.O.Box 5800, MS 1110
Albuquerque, NM 87185-5800
USA
e-mail: scistra@cs.sandia.gov
homepage: <http://www.cs.sandia.gov/~scistra/>

Pavel Pevzner
Department of Mathematics, DRB 155
University of Southern California
Los Angeles, CA 90089-1113
USA
e-mail: ppevzner@hto.usc.edu
homepage: <http://www-hto.usc.edu/people/Pevzner.html>

Ron Shamir
Department of Computer Science
School of Mathematics
Tel Aviv University
Tel Aviv 69978
ISRAEL
e-mail: shamir@math.tau.ac.il
homepage: <http://www.math.tau.ac.il/~shamir/>